

A Symmetry Breaking Model for X Chromosome Inactivation

Mario Nicodemi^a and Antonella Prisco^b

^a Dip. di Scienze Fisiche, Univ. di Napoli ‘Federico II’, INFN, Via Cintia, 80126 Napoli, Italy

^b CNR Inst. Genet. and Biophys. ‘B. Traverso’, Via P. Castellino 111, 80131 Napoli, Italy

In mammals, dosage compensation of X linked genes in female cells is achieved by inactivation of one of their two X chromosomes which is randomly chosen. The earliest steps in X-inactivation (XCI), namely the mechanism whereby cells count their X chromosomes and choose between two equivalent X, remain mysterious. Starting from the recent discovery of X chromosome colocalization at the onset of X-inactivation, we propose a Statistical Mechanics model of XCI, which is investigated by computer simulations and checked against experimental data. Our model describes how a ‘blocking factor’ complex is self-assembled and why only one is formed out of many diffusible molecules, resulting in a spontaneous symmetry breaking (SB) in the binding to two identical chromosomes. These results are used to derive a scenario of biological implications describing all current experimental evidences, e.g., the importance of colocalization.

X chromosome inactivation (XCI) is the phenomenon in female mammal early embryo cells by which one of their two X chromosomes, randomly chosen, is transcriptionally silenced, and epigenetically inherited in descendants, to equalize the dosage of X genes products with respect to males [1, 2, 3]. Crucial aspects of this chromosome-wide stochastic regulatory mechanism, necessary to survival, still elude comprehension despite being the focus of substantial interest for their important scientific and medical implication (see [1, 2, 3] and Ref.s therein). Starting from the important discovery of X colocalization during XCI establishment [4, 5], in this paper we propose a Statistical Mechanics model of the early steps of XCI.

Actually, XCI is a multistep process involving [1, 2, 3]: “counting” the number of the X chromosomes of the cell, “choice” of the inactive X, its silencing and maintenance. Silencing and maintenance start being understood: the former is induced by the action of the *Xist* gene transcript, and maintenance of the inactive state is a paradigm of epigenetic inheritance [3, 6]. Counting and choice are, instead, in many respects still mysterious, though it is well established they are controlled by yet unknown sites located within a 1 Mb region on the X, the X-chromosome-inactivation center (*Xic*), containing several genes and regulators [1, 2], such as the *Xist* gene. We also know that cells having a normal number of autosomes (non sex chromosomes) and extra copies of the X chromosome have only one active X, irrespective of the number of X’s [1, 2, 7].

This biological scenario suggests [1, 2, 3] that “controlling factors” for counting and choice derive from autosomes and interact with cis-acting regulatory sequences on the X chromosomes, whose position within the *Xic* is still unknown. Current models postulate the existence of a “blocking factor” (BF) [1, 2, 3], a complex made of X and autosomal factors, binding to the *Xic* of just one chromosome per diploid cell preventing its inactivation, as the second unprotected *Xic* in a female cell is inacti-

vated by default. Multiple factors models were proposed as well [8, 9]. These models do not account, though, for the discovery that colocalization, i.e., a physical proximity, occurs between X chromosomes at the onset of XCI, specifically in the *Xic* region, a phenomenon shown to be necessary for XCI [4, 5]. To comprehend the role of *Xic* colocalization, a description of the system is demanded that considers the influence of the spatial configuration on the interaction of the *Xic* with the BF.

The molecular nature of the blocking factor is itself still unknown: it must be a unique entity to perform its function, though, several considerations (e.g., degradation, over-production problems) exclude the possibility that it is a single protein or RNA molecule. On the other hand, if a diffusible controlling factor is produced in several copies that can statistically reach the target, asymmetric binding to two equivalent chromosomes must be explained. While the BF can be envisaged as a unique complex formed by autosomally derived molecules, why only one is formed is not understood.

We introduce a Statistical Mechanics schematic model of the diffusible “controlling factors” theory of X inactivation and we explain how a supermolecular complex can be self-assembled and why only one is formed out of many molecules, resulting in a spontaneous symmetry breaking (SB) in the binding to two identical chromosomal targets. We use, then, our new insights on the “blocking factor” to derive biologically relevant implications and depict a comprehensive scenario of experimental evidences that highlights the implications of *Xic* colocalization with respect to the kinetics of XCI.

Our model - In our model, for simplicity, we include just the essential components of the process we are interested in (see Fig.1): the two relevant proximal portions of, say, the *Xic*, where the diffusible controlling factors are assumed to bind, and a portion of space surrounding them. Such a volume includes an initially random distribution of molecular controlling factors originated by autosomes. These factors are represented by diffus-

ing particles having an affinity for their target regions on the two X chromosomes, as well as a reciprocal affinity among themselves, as they can form a complex.

In our schematic description, the two *Xic* segments are parallel, at a given distance L in some units d_0 (of the order of the unknown molecular factors size), in a volume (a cubic lattice with spacing d_0) of linear sizes $L_x = 2L$, $L_y = L$ and $L_z = L$ around them (see Fig.1). The diffusing factors randomly move from one to a nearest neighbor vertex on the lattice. On each vertex no more than one particle can be present at a given time [10].

Each particle interacts with its nearest neighbors via an effective energy E_0 . Below, we mainly discuss the case where E_0 is of the order of a ‘weak’ hydrogen bond energy, say 6 kJ/mole, which at room temperature corresponds to $E_0 = 2.4kT$ [11] (the “random walk” model is recovered if $E_0 = 0$). The probability of a particle to move to a neighboring empty site is proportional to the Arrhenius factor $r_0 \exp(-\Delta E/kT)$, where ΔE is the energy change in the move, k the Boltzmann constant and T the temperature [11, 12]. The prefactor r_0 is the reaction kinetic rate (setting the time scale here), depending on the nature of the molecular factors and of the surrounding viscous fluid (for example, $r_0 = 30\text{sec}^{-1}$ is a typical value of biochemical kinetics). Finally, since the *Xic* chromosome segments have an affinity for particles, each lattice site belonging to the chromosomes has a binding energy E_X (equal for the two X’s) with particles; for simplicity we take $E_X = 2.4kT$ too.

The idea we illustrate below is that the molecular factors interaction, E_0 , induces cluster formation (see Fig.1): when a freely diffusing particle collides with a cluster of other particles, it tends to “stick” to them, which produce cluster growth. We show that if E_0 is above a given threshold, E^* (of the order of ‘weak’ hydrogen bonds), a phase transition occurs and the many clusters eventually coalesce in a single major “complex”. Interestingly, the time, τ , to form the complex rapidly grows with the X segments distance, L , explaining the important role of X colocalization.

Computer Simulations - We studied by Monte Carlo simulations [13] the dynamics and the final state attained by the system. The size of our lattice is $L = 16d_0$ (we checked our results for L as large as $128d_0$), with periodic boundary conditions. Averages are over at least 256 runs and time is given in units of Monte Carlo lattice sweeps [13]. The fraction of particles per lattice site in the examples below (see Fig.1) is $c = 25 \cdot 10^{-3}$.

Pictures of the system state at given time slices during a typical evolution are shown in Fig.1, which compares the simple “random walk” model ($E_0 = 0$) with the present model ($E_0/kT = 2.4$). The difference between the two is apparent: in the “random walk” case, particles diffuse without forming any structure except for some binding to the ‘chromosomes’. When $E_0 = 2.4kT$, clusters of particles form, which end up in a single big

cluster covering only one of the ‘chromosomes’. This phenomenon, similar to nucleation where DNA acts as a seed [14], illustrates how the formation of a single “complex” and the spontaneous breaking of the binding symmetry between the two X chromosomes can occur. Notice that if the chromosome affinity tends to zero, $E_X \rightarrow 0$, a single cluster of particles is still eventually formed, but its binding to the X’s is unlikely.

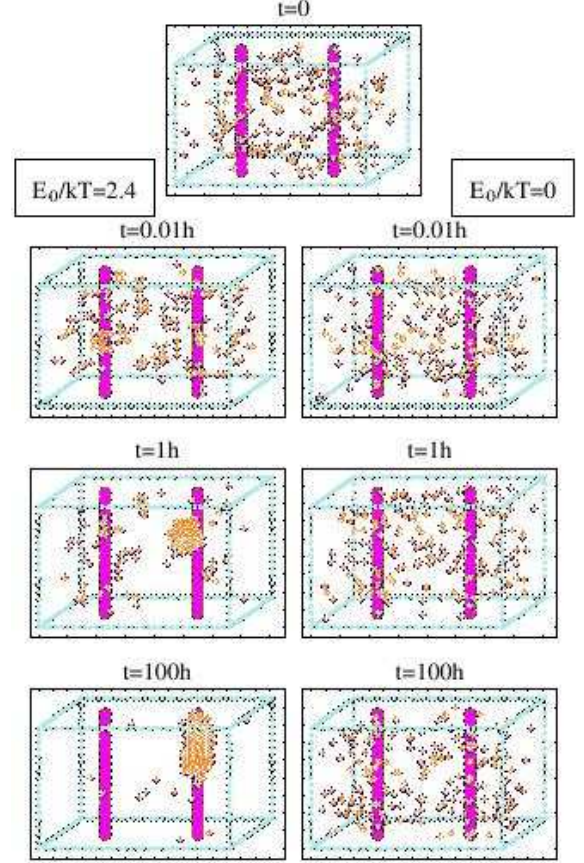


FIG. 1: We show the evolution of our particle system, around two equally binding “chromosomes”, when the effective particle interaction energy is $E_0 = 2.4kT$ (**left**) and $E_0 = 0$ (**right**), starting from the same initial random configuration at $t = 0$. When $E_0 = 0$, a “random walk” diffusion is found. The evolution is drastically different for $E_0 = 2.4kT$, where droplets of particles are formed, ending up in a single cluster covering just one of the two equivalent chromosomes and, thus, breaking their binding symmetry.

Fig.2 shows the evolution, during the same kind of run of Fig.1, of the average concentration around the chromosome on the left, ρ_l , and on the right, ρ_r ($\rho_l = N_l/V_l$ where N_l is the number of particles in a cylinder, of radius $R = 2.5d_0$ and volume $V_l = \pi R^2 L_z$, centered around the left chromosome; analogously $\rho_r = N_r/V_r$). In the “random walk” case, $E_0 = 0$, the naive expectation from the symmetry between the two chromosomes that $\rho_l(t) = \rho_r(t)$ is indeed verified at all times, t ; $\rho_l(t)$ and $\rho_r(t)$ slightly grow up to a value comparable to the

initial random one. In the $E_0/kT = 2.4$ case, ρ_l and ρ_r start from the same initial value, c , but at some point one of the two has a crisis as particles are all taken in the region of the other X whose local concentration gets one order of magnitude larger than at the beginning.

The assembling of a single factor is attained when a balance is achieved between the entropy reduction and energy gain in the process: for a given concentration of the particles, c , only when the interaction energy, E_0 , is above the phase transition line value $E^*(c)$ (broken line in the phase diagram of the right inset in Fig.2), a single major complex is formed and, thus, the original binding symmetry between the two chromosomes is broken.

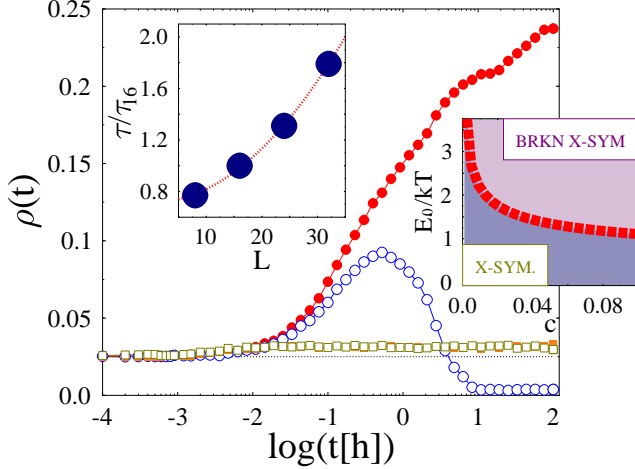


FIG. 2: The average concentration around the chromosome on the left, ρ_l , and on the right, ρ_r , of Fig.1 are shown as a function of time in a logarithmic scale. Circles refer to the case $E_0 = 2.4kT$ and squares to $E_0 = 0$ (filled symbols for ρ_l , empty symbols for ρ_r). If $E_0 = 0$, the symmetry between the two chromosomes is preserved during the evolution and $\rho_l(t) = \rho_r(t)$. When $E_0/kT = 2.4$, the symmetry is broken: after a transient, ρ_l grows one order of magnitude larger than at the starting point and $\rho_r \rightarrow 0$. **Right inset** The system phase diagram, derived by simulations, is shown in the interaction energy-concentration plane, (E_0, c) , where the broken curve fits the phase transition line. **Left inset** We plot the time, τ , to assemble the BF as a function of the distance L between the two X segments in a box of given size $L_x = l_0$, $L_y = l_0/2$ and $L_z = l_0/2$ (with $l_0 = 64d_0$, $E_X = E_0 = 2.4kT$ and $c = 25 \cdot 10^{-3}$). Specifically, we show the ratio $\tau(L)/\tau_{16}$, where τ_{16} is the reference value for $L = 16d_0$, the distance considered in the rest of the paper. The fit is: $\tau = a + bL^2$.

Discussion - The SB mechanism we illustrated can explain why only one X chromosome per cell is active: it is the only one to recruit a sufficient number of controlling factors. The time, τ , to complete the assembling of the blocking factor complex (BF) can be defined from the dynamics of the system “order parameter”, $m(t) = |\rho_l(t) - \rho_r(t)|/(\rho_l + \rho_r)$, which at long times is approximately exponential: $m(t) \propto \exp(-t/\tau)$. Alike Brownian processes, τ increases as the square of the X segment distance, L : $\tau \sim L^2$ (see left inset of Fig.2). This suggests that only when the X’s colocalize the BF

can be assembled in a time short enough to be useful on biological time scales.

Differences in the affinities of the X, E_X , induced for instance by deletions of binding sites on one of them, result in a decisive breaking of the symmetry in particle binding: the X with less affinity remaining “naked” and, thus, unprotected from inactivation. This explains biased XCI in embryonic tissues resulting from allelic differences in *Xic* sequences, (e.g., in the *Xce* locus [15] or other regions [1, 2, 3]). XCI in female cells and lack of XCI in male cells could derive from a similar mechanism. With respect to XCI in polyploid cells, in our simulations the higher is c the larger the probability to have, at intermediate stages, two clusters, bound to two X, large enough to act as BF’s. This could describe the stochastic nature of the ‘X chromosome/autosome ratio effect’ [1, 2, 3].

The SB model also rationalizes recent important deletion experiments across the *Xic* region, known to affect choice and counting (see [8, 9, 16, 17, 18, 19, 20] and Ref.s therein). We summarize here, in particular, the phenotype of three deletions, which were instrumental in defining the role of the region 3’ to *Xist* in counting and choice, namely $\Delta 65\text{kb}$ [16], *Tsix* $^{\Delta CpG}$ [9] and *Xite* $^{\Delta L}$ [19].

The $\Delta 65\text{kb}$ deletion removes 65kb of DNA in the *Xic* region relevant to the chromosome activation [16]. $\Delta 65\text{kb}$ causes non-random inactivation of the deleted X in heterozygous XX cells[16], and X inactivation in XY cells[18]. The explanation from BF models is that the $\Delta 65\text{kb}$ deletion removes the binding sites for the blocking factor and the complex cannot bind the X any more. Interestingly, the X chromosome bearing the deletion is not active, not even in male cells.

The behavior of male cells is, however, drastically different in the case of shorter deletions. The analysis of two smaller non-overlapping deletions within the above mentioned $\Delta 65\text{kb}$ sequence, namely the *Tsix* $^{\Delta CpG}$ deletion, removing the *Tsix* promoter, and the *Xite* $^{\Delta L}$ mutation, removing *Xite*, added further important information. In heterozygous XX cells, the *Tsix* $^{\Delta CpG}$ deletion causes non-random inactivation of the deleted X, whereas in XY cells the *Tsix* $^{\Delta CpG}$ deleted X remains active[9]. The *Xite* $^{\Delta L}$ phenotype is analogous [19]. Finally, in homozygous *Tsix* $^{\Delta CpG}$ XX mutants the choice of the active X is still random[21], but, importantly, in a fraction of cells both X’s are inactivated (“chaotic counting” [8]).

As usual single BF models cannot explain these results [8], the simulations with the SB version we propose (summarized in Fig.3) allow a fresh look at the *Tsix* $^{\Delta CpG}$ and *Xite* $^{\Delta L}$ data. In our model, the blocking factor is a cluster of transacting factors which can bind several sites on a chromosome at the same time. The *Tsix* $^{\Delta CpG}$ deletion reduces the total blocking factor/chromosome affinity. In our model, the difference in the affinity, E_X , of the wild type and of the deleted chromosome explains, as described before, why choice is skewed in the heterozygous XX cells (see Fig.3A,B). At

variance with the results from the longer $\Delta 65\text{kb}$ deletion, however, in the case of the smaller $Tsix^{\Delta CpG}$ and $Xite^{\Delta L}$ deletions, the mutated X remains active in XY cells. This is easily understood within the SB model: if binding sites are found in both the regions deleted by $Tsix^{\Delta CpG}$ and $Xite^{\Delta L}$, then each mutation will reduce the affinity of the chromosome for the blocking factor, though neither deletion will fully erase the overall affinity. Thus, in XY cells the blocking factor can still bind the deleted X chromosome (see Fig.3C,D), since there is no competing wild type X. Random choice in homozygous XX is explained as before (see Fig.3E). “Chaotic counting” [8] in homozygous deleted mutants derives from an analogous mechanism: deletions significantly reduces the total X-chromosome/particles affinity (see Fig.3F,G) and, in a fraction of cells, the blocking factor doesn’t bind at all.

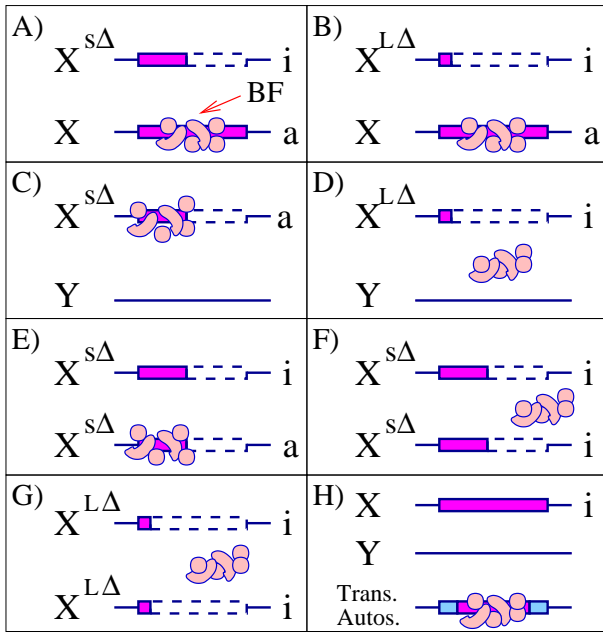


FIG. 3: A pictorial summary of SB model results. Panels A,B) consider heterozygous XX cells with either a comparatively long deletion in the Xic region ($X^{L\Delta}$), say $\Delta 65\text{kb}$ [16], or a shorter deletion ($X^{s\Delta}$), say $Tsix^{\Delta CpG}$ [9]. The mutated X, having a reduced overall affinity for the blocking factor (BF), loses on average the competition for it (*skewed inactivation*). C,D) In XY cells, the effects of $X^{s\Delta}$ and $X^{L\Delta}$ can be quite different: $X^{L\Delta}$ being unable to bind the BF. E,F) In homozygous XX cells $X^{s\Delta}$, with a reduced affinity, succeeds in binding the BF only in a fraction of cases (“chaotic counting” [8]), whereas G) $X^{L\Delta}$ is unable to bind BF. H) In XY cells, transgenic autosomes with long enough Xic insertions can bind the BF and the X is inactivated in a fraction of cases.

Transgene insertions into autosomes have also been analyzed [15, 22, 23]. When long Xic transgenes are introduced, in multiple copies [15], into autosomes of male ES

cells, inactivation of the single X occurs in a fraction of the cells [22, 23]. In our view the mutated autosomes can bind the BF and compete with the X for it, leaving the real X chromosome prone to inactivation (see Fig.3H).

The simple version of the SB model here discussed considers only a single kind of BF complex, though the model could easily accommodate more than one (in a Potts-like variant), as recently proposed in [8]. It does not either imply that only one kind of soluble factors is involved.

Summarizing, the SB regulatory mechanism we propose describes the self-assembling of the blocking factor from diffusible DNA binding molecules and explains why only one is formed, i.e., the binding symmetry of the two equivalent X’s is broken. The emerging picture of its properties helps in delineating a scenario of biological implications reconciling within a single framework the existing experimental evidences (e.g., X colocalization). In our model “counting” and “choice” are no longer distinct phenomena: they are regulated by the SB stochastic mechanism where time is an important parameter. More generally, the simplicity and robustness of the SB mechanism, illustrated here for XCI, suggest it can be widely used in random monoallelic expression processes [24].

-
- [1] Avner, P., and Heard, E. (2001) *Nat. Rev. Genet.* **2**, 59.
 - [2] Lucchesi, J.C., Kelly, W.G. and Panning, B. (2005) *Annu. Rev. Genet.* **39**, 615.
 - [3] Chow, J.C., Yen, Z., Ziesche, S.M., Brown, C.J. (2005) *Annu. Rev. Genomics and Hum. Genet.* **6**, 69-92.
 - [4] Na Xu, Tsai, C.-L., Lee, J.T. (2006) *Science* **311**, 1149.
 - [5] Bacher, C.P., et al. (2006) *Nat. Cell Biol. Letters* **8**, 293.
 - [6] de Napoles M., et al. (2004) *Dev. Cell.* **7**, 663-76.
 - [7] Gartler, S.M., Riggs, A.D. (1983) *Ann. Rev. Gen.* **17**, 155.
 - [8] Lee, J.T. (2005) *Science* **309**, 768-71.
 - [9] Lee, J.T., Lu, N. (1999) *Cell.* **99**, 47-57.
 - [10] This version is a generalization of the Ising model [12].
 - [11] Watson, J.D., et al. “*Molecular Biology of the Gene*”, Benjamin Cummings (2003).
 - [12] Stanley, H.E. “*Introduction to Phase Transitions and Critical Phenomena*”, Clarendon Press (1971).
 - [13] Binder, K. (1997) *Rep. Prog. Phys.* **60**, 487.
 - [14] Cacciuto, A., et al. (2004) *Nature* **428**, 404.
 - [15] Heard, E., et al. (1999) *Mol. Cell. Biol.* **19**, 3156-3166.
 - [16] Clerc, P., and Avner, P. (1998) *Nat. Genet.* **19**, 249-53.
 - [17] Clerc, P., Avner, P. (2003) *Sem. Cell. Dev. Biol.* **14**, 85.
 - [18] Morey, C., et al. (2004) *EMBO J.* **23**, 594-604.
 - [19] Ogawa, Y., Lee, J.T. (2003) *Mol. Cell.* **11**, 731-43.
 - [20] Morey, C., et al. (2001) *Hum. Mol. Genet.* **10**, 1403-11.
 - [21] Lee, J.T. (2002) *Nat. Genet.* **32**, 195-200.
 - [22] Lee, J.T., Strauss, W.M., Dausman, J.A., Jaenisch, R. (1996) *Cell.* **86**, 83-94.
 - [23] Herzing, L.B., et al. (1997) *Nature* **386**, 272-5.
 - [24] Singh, N., et al. (2003) *Nat. Genet.* **33**, 339-341.